

TARTU ÜLIKOOL
FILOSOOFIATEADUSKOND
EESTI JA ÜLDKEELETEADUSE INSTITUUT
EESTI KEELE OSAKOND

Birgitta Ojamaa

TARTU ÜLIKOOLI ÜLIÕPILASTE TAGASISIDE HOIAKUTE ANALÜÜS

Bakalaureusetöö

Juhendaja Sven Laur

Tartu 2014

SISUKORD

SISSEJUHATUS	3
1. HOIAKUTE ANALÜÜS	6
1.1. Meetodid	8
1.2. Tunnete väljendamine	10
1.2. Tunnete väljendamine eesti keeles.....	12
2. TARTU ÜLIKOOLI ÜLIÕPILASTE TAGASISIDE HOIAKUTE ANALÜÜS	14
2.1. Tulemused.....	18
KOKKUVÕTE	22
KIRJANDUS.....	24
SENTIMENT ANALYSIS ON UNIVERSITY OF TARTU STUDENT FEEDBACK.	
SUMMARY	26
LISAD.....	27
Lisa 1. Tundeleksikon	27

SISSEJUHATUS

Hoiakute analüüs, tuntud ka kui tundmusanalüüs (ingl *sentiment analysis*, *opinion mining*) või tehisintellekti valdkonnas afektirehkendus (*affective computing*), võimaldab informaatika- ning lingvistikateadmiste põhjal luua inimese emotsiooni, hinnanguid ning suhtumist analüüsivat programmi (Vainik 2010: 330). Hoiakud, mida uuritakse, seostuvad toodete, teenuste, organisatsioonide, isikute, probleemide, sündmuste ning teemade täpsemate omadustega. Analüüsitavat materjali kogutakse Internetist sotsiaalmeedia vahendusel.

Vaatamata sellele, et paljud internetilehed annavad kasutajatele võimaluse hinnata tooteid ning teenuseid näiteks punktide või *like*-nuppude abil, ei anna niisugused kokkuvõtvad hinnangud detailset ülevaadet sellest, mille kohta inimesed hästi või halvasti arvasid. Seetõttu tuleb kasuks uurida arvamustekste. (Liu 2012: 2)

Erinevalt fakti konstateerivatest lausetest on arvamused ning suhtumislauseid subjektiivsed. Seega on tähtsam uurida suurt hulka arvamusi paljudelt inimestelt kui ühte arvamust ühelt inimeselt, et teha paremaid järeldusi.

Kuna tundeid, nagu ka arvamusi, on palju ning kõigi emotsioonide väljendamist kirjalikus vormis pole uuritud, ei tehta analüüsi igast tekstides esinevast tundest või arvamusest. Peamiselt keskendutakse vaid positiivsete ja negatiivsete tunnete määratlemisele (Liu 2012: 1).

Kuigi keeletehnoloogia ning lingvistika aladel on kaua aega koostööd tehtud, hakati inimeste arvamusi ja hinnanguid alles 2000. aastatel rohkem uurima. Internetis leviva vabalt kättesaadava informatsiooni tõttu on hoiakute analüüs nüüd populaarsem. Termin *sentiment analysis* ilmus esiteks Nasukawa ja Yi uurimuses „Sentiment analysis: Capturing favorability using natural language processing.” 2003. aastal, termin *opinion*

mining samuti 2003. aastal Dave'si jt uurimuses „Mining the peanut gallery: Opinion extraction and semantic classification of product reviews”. (Liu 2012: 1)

Hoiakute analüüsi suurimad probleemid tulenevad keele eripäradest. Põhilisteks probleemideks on konteksti tuvastamine, mittenormaallausete määramine, sarkastilised laused ning pealtnäha objektiivsetes lausetes arvamuste tuvastamine.

Hoiakute analüüsi uurimise põhiliseks põhjuseks on erinevate rakenduste loomine (Liu 2012: 3). Rakendused keskenduvad nii erinevatele tekstiliikidele, näiteks Twitteri tekstide analüüsimine (nt Tweet Sentiment Visualization App, Sentiment140) või ajaveebitekstide analüüsimine (nt Repustate Sentiment Analysis), kui ka arvustuste eriliikidele, näiteks käsitletakse arvamusi teenindusest või poliitkampaniatest teistmoodi kui toodete arvustusi.

Tüüpiline rakendus märgendab või eristab arvamustekstist sõnad, millel on kas positiivne või negatiivne emotsionaalne laeng. Laengutega sõnade puhul arvestatakse, millist gruppi esineb rohkem ning selle järgi määrataksegi arvamuse polaarsus – kas negatiivne või positiivne suhe lugejatele. Seda saab näha joonisel 1, kus Eesti Keele Instituudi emotsioonidetektor on analüüsinud arvamusteksti (Emotsioonidetektor). Negatiivset hoiakut väljendavad sõnad on märgistatud sinise taustaga, nagu joonisel on näha. Positiivsed sõnad aga tuuakse välja rohelisel taustal. Pärast sõnade väljatoomist tekstis leiab rakendus tekstis üldise hinnangu, mis selle tekstilõigu puhul oli negatiivne.

Ministri sõnul peab tõlked garanteerima ravimifirma. See tähendab, et kulud tõlkimise ja info levitamise eest kannab ravimifirma. Kuidas on aga seadustega? Seaduste tõlkimine vene keelde on Rõivase sõnul raske ja kulukas ning riik sellega tegelema ei peaks.

Hinnang: negatiivne

Joonis 1. Eesti Keele Instituudi emotsioonidetektori analüüs. Tekst pärineb arvamusloost „Seadused võiks avaldada ka vene keeles“. (Himma 2014)

Eestis on hoiakute analüüsis tegeletud vaid uudistekommentaaridega (Marran 2012) ning ehitatud emotsioonidetektor, mis tegeleb tekstilõikudest emotsionaalse info leidmise ning määratlemisega (Eesti Keele Instituudi emotsioonidetektor). Vähese huvi põhjustajatenä Eestis võib välja tuua eesti keele keerukuse (näiteks käänete-pöörete tõttu) ning ka vähese emotsioonisõnavara uurimise. Peamiselt on emotsioonisõnavara Eestis uurinud vaid Ene Vainik. Kindlasti on mõjutajaks ka korpuste vähene maht ning keeleülest hoiakute analüüsi abistavate tööriistade, näiteks tundesõnastike, puudumine.

Selle töö eesmärk on anda ülevaade hoiakute analüüsist Eestis ning mujal maailmas ning luua õppeinfosüsteemist saadud Tartu Ülikooli üliõpilaste tagasiside korpusele hoiakuid analüüsiv süsteem. Lisaks proovitakse hinnata, kui hästi süsteem suudab eestikeelsetest lausetest hoiakuid leida.

Töö koosneb kahest peatükist. Esimeses peatükis selgitatakse, mis on hoiakute analüüs, kuidas ning milleks hoiakute analüüsi rakendada. Lisaks uuritakse, millised on tüüpilised keelest tulenevad takistused hoiakute analüüsi süsteemidel. Teine osa keskendub rakendusele, mis analüüsib Tartu Ülikooli õpilaste tagasisides esinevate hoiakute positiivsust ja negatiivsust.

1. HOIAKUTE ANALÜÜS

Hoiakute analüüsist üldist pilti anda on keeruline, kuna analüüsiva süsteemi loomiseks võib kasutada erinevaid viise ning erinevaid abimaterjale; sellest on võimalik pikemalt lugeda peatükis 1.1. Peamiselt jagunevad hoiakute analüüsi süsteemid kahte rühma: reeglipõhised ning statistikal põhinevad süsteemid.

Hoiakute analüüsiks luuakse sobiv süsteem olenevalt tekstidest, mida analüüsida soovitakse, ning kasutatavatest abimaterjalidest. Üldiselt kasutatakse abiks emotsioonileksikone, meelsusvihjeid ning erinevaid reegleid. Emotsioonileksikonid näitavad, kas sõna on positiivne või negatiivne. Neist saab pikemalt lugeda peatükis 2. Meelsusvihjeid viitavad sellele, et lauses on mõni arvamus. See on kasulik, kui leksikoni abil arvamust leida on keeruline. Reeglite eesmärk on süsteemile vihjeid anda, kuidas sõnad lauses üksteist mõjutavad, et vigu leida, ning otsida ka neutraalseid ja nii positiivse kui negatiivse väärtusega sõnu (Wilson jt 2005: 348).

Üheks hoiakute analüüsi uurimise põhjuseks on rakenduste loomise võimalus. Hoiakute analüüsi rakendusi saaks kasutada erinevates valdkondades. Avalikkuse arvamuste kogumine on tulnud kasuks turundusele, suhtekorraldusele ning poliitiliste kampaaniate korraldajatele, kuna arvamused mõjutavad ka teiste inimeste käitumist ning võivad mõjutada teatud firmade ning organisatsioonide suhtumist tarbijatesse. (Liu 2012: 3)

Rakendusi võib olla erinevaid; tekib vajadus eraldi reeglite järele, olenevalt rakenduse eesmärgist, tekstide keelest, allkeelest ja tekstide pikkusest.

Hoiakute analüüs võiks asendada varem arvamuste uurimiseks kasutatud uuringud, küsitlused ning fookusgrupi-uuringud. Paljudel suurfirmadel, näiteks Microsoftil (*Microsoft Dynamics*) või Google'il (*Google Prediction API*), on projektid hoiakute analüüsi arendamiseks.

Hoiakute analüüsi süsteeme on enamasti loodud inglise keelele, mille suured märgendatud korpused, tesaurused, WordNet ja tundesõnastikud hõlbustavad rakenduste ehitamist ja testimist. Abiks on ka inim- ja rahalised ressursid. Ka vene keelele on loodud rohkem toimivaid lahendusi. (Vainik 2010: 330)

Väiksematele keeltele on vähe lahendusi loodud. Seetõttu on keelteülestes tööriistade ning korpuste järele vajadus tungiv. Selle mõjuriks on esiteks erinevates keeltes hoiakute süsteemide hõlpsam loomine, teiseks suurfirmade vajadus muukeelsete klientide ning ingliskeelsete klientide arvamusi kõrvutada (Liu 2012: 34). Keelteülestes tööriistade ning materjalide loomiseks kasutatakse tihti masinõpet, mõnikord kombineerides seda muude reeglite või näiteks WordNetiga (Liu 2012: 34–35). Masinõppe meetodite abil on võimalik märgendatud tekstidest tuletada peidetud reegleid ning WordNetis kättesaadavad sünonüümi-antonüümi ning hierarhiaseid saab kasutada sõnastiku laiendamiseks. Keelteülestes tööriistade arendamiseks võib kasutada ka paralleelkorpusi (Liu 2012: 45).

Eesti keeles on hoiakute analüüsi kohta kirjutatud üks bakalaureusetöö: Siim-Toomas Marrani „Sentimentaalne analüüs eestikeelse peavoolumeedia veebiartiklite kommentaaride baasil“ (2012). Selle töö eesmärgiks oli ehitada hoiakute analüüsi rakendus uudistetekstide analüüsimiseks sõnastiku ning statistilise klassifitseerija põhjal (Marran 2012).

Eesti Keele Instituudis on riikliku programmi „Eesti keele tehnoloogia 2011–2017“ projekti „Kõne ja teksti emotsionaalsuse statistilised mudelid“ raames loodud emotsioonidetektor. See tuvastab kirjaliku teksti lõikude positiivsuse, negatiivsuse ja neutraalsuse sõnastiku põhjal. Detektor kasutab ka statistilist klassifitseerijat. Emotsioonidetektori põhieesmärgiks on edendada inimeste ning masinate vahelist suhtlust, kuid seegi võib olla komponendiks hinnangute ning arvamuste uuringutes. (KT 2011–2017)

Lisaks on tundesõnu ning nende väljendamist tekstis uurinud ka emotsioonidetektori loomisega seotud Ene Vainik. Selle töö jaoks on tähtsamad Vainiku artiklid ning magistritöö emotsioonikategooriast ning tundetoonidest eesti keeles. (Vainik 2002, 2012)

Peale emotsiooni tekstilise väljenduse on tegeletud ka emotsiooni tuvastamisega kõnes ning koostatud emotsionaalse kõne korpus (Pajupuu 2012; Vainik 2010; Eesti emotsionaalse kõne korpus).

1.1. Meetodid

On nii reeglipõhiseid kui automaatseid hoiakute analüüsi süsteeme. Süsteemide erinevus võib oleneda ka tekstiliigist. Erinevaid tekstiliigid vajavad erinevaid tööriistu: netikeele analüüsimiseks on näiteks vaja paindlikumat sõnastikku, ilukirjandusliku teksti jaoks rohkem reegleid.

Reeglipõhises hoiakute analüüsis kirjutatakse süsteemile ette keele ning emotsioonide väljendamise reeglid ning süsteem hakkab reeglite järgi ning sõnastiku abil teksti märgendama ja teksti polaarsust määrama. Võib otsida ka viitefraase või -sõnu, mis tundelauses tüüpiliselt esinevad, näiteks *tundma*, *arvama*, *jõudma järeldusele* jne (Liu 2012: 40).

Statistilises hoiakute analüüsi süsteemis pole käsitsi kirjutatud reegleid, süsteem võrdleb korpusi ennustusmudelitega ning pärast seda proovib kõige sobivama mudeli abil otsustada, kuidas sõnu ja lauseid märgendada. Mudelid koosnevad märgendamise reeglitest või hierarhilistest reeglistikest, mis võivad olla matemaatilisemad ning keerukamad kui reeglid reeglipõhistes süsteemides.

Emotsioonisõnade ja nende teemade või sihtmärkide ehk aspektide (*aspect*) määramiseks on meetod nimega *double propagation*, kus paari tundesõna ja aspekti

abil leitakse üha enam sõnu ja aspekte. Kuna emotsionaalse laenguga sõnad ning see, millele need viitavad, on tihti koos, leitakse sõltuvussuhted, mis on üsna sarnased sõltuvusgrammatikale. (Liu 2012: 68–69)

Sõnastiku genereerimiseks saab ka otsida paar tundesõna alguseks, läbi WordNeti lisada sõnastikku erinevad sünonüümid ja antonüümid ning leitud sõnade sünonüümid ja antonüümid. (Liu 2012: 80)

Enamik statistilisi süsteeme on treenitud märgendatud korpuste abil (*supervised learning*). Märgendatud korpustega statistilised süsteemid kasutavad sõnade jaotamiseks meetodeid, mis põhinevad traditsionaalsetel märgendatud korpuste masinõppe mudelitel nagu *Support Vector Machine* (SVM) või *naïve Bayes* (NB). (Liu 2012: 24)

Naïve Bayes on ilmselt mudelitest enim kasutatav, kuna see on suhteliselt lihtne meetod ning ei vaja palju ressursse. NBs arvutatakse välja sõna õiges kontekstis esinemise tõenäosus, sõna esinemise arv samades kontekstides korrutatakse kontekstide ja sõna esinemise arvu jagatisega. See on kerge viis sõnu klassifitseerida. (Weiss jt 2010: 55)

On ka statistilisi süsteeme, mis pole treenitud märgendatud korpuste peal (*unsupervised learning*). Neis kasutatakse tundesõnade leksikoni ning otsitakse teatud fraase lauseümbruse põhjal (Liu 2012: 28–29). Lisaks on statistiliseks hoiakute analüüsiks kasutatud märgendatud ja märgendamata korpuste kooslust (*semi-supervised learning*). Väiksema märgendatud korpuse informatsiooni kasutades proovib süsteem traditsiooniliselt suuremat märgendamata korpust analüüsida (Liu 2012: 30, 32).

Statistilisel analüüsil on ka halbu omadusi. Näiteks võib statistilise klassifitseeriija treenimine hoiakute analüüsiks osutuda raskeks, kuna on vaja palju märgendatud arvamustekste.

Lisaks võib esineda probleeme, kuna vaid sõnastiku abiga leiavad statistilised süsteemid üles umbes 60% sõnadest. Ülejäänud sõnad erinevad teistest märgatavalt, pole tavakasutuses või on nii mitmeste tähendustega, et tavaline statistiline süsteem suuda neid leida. (Liu 2012: 77) Näiteks eesti keeles fraas „vääriti käituma“, kus *vääriti* on omadussõna *väär* suhteliselt vähekasutatav vorm, seega statistiline süsteem võib leida, et *vääriti* on pigem tegevussõna *väärima* verbivorm või hoopis seosetu väljendverb.

Lühemate tekstide analüüsimiseks kasutatakse vaid sõnastikku, mõnikord ka mõnda reeglit. Seda nimetatakse *grep*- (*globally search regular expression and print*) meetodiks. Ehk siis otsitakse tekstist üles vaid sobivad sõnad või sõnajupid ning leitakse, mis hinnangut lause väljendab. Näiteks lausetel „*toode on väga hea*“, „*suurepärane mu arust, ei kahetse*“ ja „*suht jama*“ polegi vaja reegleid keerulisemate lausekonstruktsioonide analüüsiks. Selle bakalaureusetöö praktilises osas kasutatigi peamiselt seda varianti, kuna üldjuhul polnud arvamused eriti pikad.

Pikemates tekstides on rohkem võimalusi. Kui kasutatakse statistilist analüüsi, selgub, et SVM mudelil põhinev masinõpe on parem, kui NB. Lühematel lõigetel toimib paremini NB-tüüpi masinõpe. (Wang, Manning 2012: 92)

1.2. Tunnete väljendamine

Lisaks rakenduste loomisele on oluline uurida hoiakute avaldumise analüüsi seoses kirja-, aga ka netikeele omadustega, näiteks kontekstitundlikkus, toon ja subjektiivse teksti eristamine objektiivsest.

Suurte korpuste puudumise tõttu ning ka tulenevalt eesti keele vormirohkusest ning vabast sõnajärjest võivad hoiakute analüüsi tegemisel tekkida probleemid, mida mõnes teises keeles ei teki. Arvamustekstide uurimise arengut takistavad ka samavormilised, kuid erineva tähendusega sõnad. Neist probleemidest on pikemalt kirjutatud alampeatükis 1.2.1.

Hoiakute analüüsis on palju uurimisküsimusi seoses keelest tulenevate eripäradega. Bing Liu (2012: 5–6) järgi toon välja keelelised probleemid, mida jätkuvalt uuritakse.

1. Positiivsel või negatiivsel tundesõnal võib erinevas kontekstis olla vastupidine tähendus. Näiteks sõna *jube* väljendab lauses „*see film oli väga jube, režissööri ilmselt polnud filmimise ajal kohal*” negatiivset arvamust, aga lauses „*see rula on jube äge*” täiendab see positiivset arvamust.

2. Tundesõnade olemasolu lauses ei näita tingimata suhtumist. Küsilauseid ning konditsionaallauseid on vastuolulised lausetüübid, milles tundesõnade mõju on raske hinnata. Näiteks laused „*kui lähen kinno, vaatan, kui hea see film on*” või „*kui läksin kinno, vaatasin seda head filmi*” on küll sarnased ja sisaldavad positiivset sõna *hea*, aga lausetest esimene on neutraalne.

3. Väga raske on analüüsida sarkastilisi lauseid. Automaatne süsteem ilmselt ei tunneks ära negatiivset lauset „*pesumasin töötab nii hästi, et pesu on uusi plekke täis*”. Praegu tehtud uurimuste põhjal saaks osa sarkastilisi lauseid teatud lausestus- ning kirjavahemärgimalli järgi ülejäänud tekstist eraldada, kuid mitte eriti täpselt – näiteks on neis lausetes palju hüüumärke, küsimärke, jutumärke ning suurtähti või suures kirjas sõnu (Liu 2012: 44).

4. Paljud tundesõnadeta laused väljendavad siiski arvamusi. Neist lausetest on paljud fakte sisaldavad objektiivsed laused. Laused „*see õppejõud annab vähe juhiseid kodutöö tegemiseks*” või „*see kustukumm jätab palju tolmu paberile*” on objektiivsed, kuid väljendavad mõlemad negatiivseid arvamusi.

Lisaks on sotsiaalmeedia eripäraks see, et anonüümsus lubab varjatud plaanide ning kuritahtlike kavatsustega inimestel jätta muljet, et nad on kõrvalseisvad isikud, ja postitada võltse arvamusi reklaamiks või millegi laimamiseks. Selliseid inimesi kutsutakse arvamusspämmeriteks (*opinion spammers*) ning nende tegevust arvamuste

spämmimiseks (*opinion spamming*). Need isikud levitavad võltse arvamusi arvustustes ning foorumipostitustes. On ka äriettevõtteid, mis kirjutavad oma klientide jaoks võltse arvustusi ning veebilogi sissekandeid. Tähtis on selliseid spämmimistegevusi avastada, et analüüsitud materjal oleks tõene. Spämmi tuvastamine erineb positiivsete ning negatiivsete arvamuste leidmisest, kuna spämmi tuvastamiseks tuleb uurida inimeste postitusharjumusi. (Liu 2012: 7)

1.2. Tunnete väljendamine eesti keeles

Eesti keeles on uuritud, et sõnavaras avalduvad emotsioonid näiteks emotsiooninimetustena, piltlike kirjeldustena ning vande- ja kirumissõnadena. Emotsionaalsele laetusele osutavad terved sõnaklassid nagu hüüdsõnad või rõhumäärsõnad. Morfoloogias avaldub emotsionaalne suhtumine vähendavate ja halvustavate sufiksita ning kliitikute kasutamisel (nt *-ke(ne)* või *-ard* lõpus sõnades *poliitikukene* või *joomard*). Süntaksis on spetsiaalsed lausetüübid, nn kogeja- või eksklamatiivlause, mille kasutamisel võivad toimuda muutused tavapärasest sõnajärjest ning konteksti mõistmist muuta. (Vainik 2010: 329)

Emotsionaalsete reaktsioonide tõlgendamine varieerub kultuuriliselt, osa emotsioonimõistetest ei kattu eri kultuurides ning keeltes üksüheselt. Ei saa eeldada, et mujal loodud emotsioonituvastaja suudaks täpselt tuvastada eestlaste emotsioone. Lisaks on seos konkreetse keelega ja selle kirjaviisiga, mis välistab teiste keelte ja kultuuride kogemuse automaatse ülevõtmise. Emotsioonide kirjutatud keeles väljendumise põhiprintsiibid aga võivad olla sarnased. (Vainik 2010: 330)

Peamiselt mõjutab uurimise käiku suurte korpuste puudus, milles oleks näiteid erinevatest tekstivaldkondadest. Eesti keeles on tekste vähem, kui keeltes, millel on rohkem rääkijaid. Kui mõnes keeles võib vabalt kättesaadavate korpuste põhjal pea igas valdkonnas töötava automaatse analüsaatori välja õpetada, siis eesti keeles piirduvad korpused vaid kõige paremini kättesaadavate tekstidega (näiteks on ilukirjanduslikku

teksti märgatavalt vähem) ning seejuures on korpuste maht silmnähtavalt väiksem. Suur osa korpustest pole ka morfoloogiliselt analüüsitud.

Eesti keele ülesehitus kõigi käänete-pöörete ning tüvemuutustega nõuab automaatselt töötlemiseks teksti morfoloogilist analüüsi. Sõnatüve ja sõnaliiki määramata poleks eesti keeles hoiakute analüüs võimalik. Tekib ka mitmeti mõistetavate sõnade analüüsimise probleem. Eriti takistab uurimist sama vormiga, kuid erineva tähendusega või mitmetähenduslike sõnade esinemine eripärastes kontekstides. Vabana sõnajärje tõttu esineb vähem lauseid, mida saaks reegli või mudeli aluseks võtta, mille alusel sõna tähendusi eristada.

Ka on eesti keele ülesehituse tõttu raskendatud automaatselt sõnapuude või sõltuvussuhete koostamine, mis aitaks määrata süntaktilisi rolle ning teha hoiakute analüüsi palju kergemaks. Suhteliselt vähe on tegeletud uue meedia tekstide süntaktilise analüüsiga, mis kirjavigade, slängi ja lühendite tõttu vajaks erinevat töötlemist (EKKT 2006–2010).

2. TARTU ÜLIKOOLI ÜLIÕPILASTE TAGASISIDE HOIAKUTE ANALÜÜS

Kasutan süsteemi ehitamiseks korpust, mis koosneb Tartu Ülikooli üliõpilaste tagasisidest ainetele õppeinfosüsteemis (ÕISis). Iga semestri lõpus peaks ÕISi kasutavad üliõpilased andma vähemalt neljale ainele tagasisidet ning hindama ka õppejõudu, seega pole analüüsitavast materjalist puudust. Üliõpilaste tagasiside on saadud anonüümselt ning seda ei tohi autor jagada.

Json-faili kujul andmebaas hõlmab vastuseid küsimusele „Mida Te ütleksite selle õppeaine kohta tulevastele õppijatele?“, mis on ka sõnedeks tehtud ning sõnesid morfoloogiliselt analüüsitud. See tähendab, et sõnede leitud ka võimalikud lemmavormid ja nende võimalik sõnaliik, kääne või pööre jms grammatiline informatsioon. Korpuses on kokku 24 762 teksti. Korpuses on nii eesti, vene kui ingliskeelset tagasisidet. Osa tagasisidet on lihtsalt suvalised tähekombinatsioonid või emotikonid. Sorteeritakse välja vaid tekstid, mis on üle üheksa tähemärgi pikad, et analüüsida täislauseid.

Süsteemi loomiseks kasutatakse Pythoni programmeerimiskeelt (versioon 2.7). Pythoni kasutamise eeliseks on, et tarkvara on tasuta, võrdlemisi kasutajasõbralik ning sellel on palju toiminguid sõnade analüüsimiseks (nt regulaararvaldised).

Eestikeelset tundesõnade sõnastikku koostati kasutades peamiselt sünonüümisõnastikku ning ka korpuses kõige sagedamini esinevate omadussõnade loendit. Kuigi oleks võinud kasutada Emotsioonidetektori sõnastikku, pole see avalikult kättesaadav ning sõnade valik on erinev. Sõnade valik põhines sellel, kuidas on võimalik väljendada positiivset või negatiivset hoiakut ainete, õppejõudude ja kodutööde suhtes. Sõnastiku sõnadel pole kaalu – sõnad pole nõrgema või tugevama tundeseose tõttu eraldi märgendatud, kuigi see tuleks kindlasti ka kasuks. Kokku tuli analüüsi teostamiseks leksikoni 419 sõna. Leksikoni positiivsed ja negatiivsed sõnad on välja toodud lisas 1.

Regulaaravaldiste abil prooviti leida eitust sisaldavaid arvamusi; üldiseid lauseid, milles võiks olla arvamus; ning lauseid, mis sisaldavad soovitusi. Lisaks taheti välja tuua konditsionaallaused, kuid see polnud ajapiirangute tõttu võimalik. Eituse otsimiseks kasutati sõnu „pole“, „mitte“ ning „ei“. Arvamuslausete otsimiseks kasutati fraaside „ma arvan, et“ ning „minu arvamus on, et“ erinevaid vorme. Võimalikeks soovituslauseteks märgistati laused, milles oli tegusõna mitmuse käsuvorm (nt „*tehke*“) või infinitiivne vorm („*teha*“). Need vormid tuuakse esile, et võimalikke soovitustekste esile tuua. Soovitustekstideski võib ju esineda tundesõnastikus esinevaid sõnu, näiteks „*tehke kodutöid, need on aine läbimiseks vajalikud*“, „*soovitan praktikumides käia, muidu kukute läbi*“ (leksikonis esinevad sõnad allajoonitud).

Eesmärgiks on hoiakute analüüsi süsteem, mis suudab sõnastiku abil eristada tekstidest negatiivsed ning positiivsed sõnad või fraasid. Nende alusel suudab süsteem otsustada, kas tegu on positiivse või negatiivse tekstiga. Leitakse sõnad ning lemmad, mis on sõnastikus. Seejärel antakse neile sõnastikuväärtuse põhjal hinnang. Eituse esinemisel pööratakse hinnang ümber – negatiivsest saab positiivne, positiivsest negatiivne. Kuigi tekstis ei pruugi olla tundesõnastikus esinevaid sõnu, võib arvamusi leida ka kasutades arvamuste leidmise viidet. Kui selline viide esineb, leiab programm sellele järgneva fraasi.

Lõpuks annab programm ka üldhinnangu, mis liiki sõnu leiti tekstist rohkem, kas positiivseid või negatiivseid. Lisaks võib programm anda ka hinnangud „neutraalne või arvamusetu tekst“, arvamussõnade ja viidete puudumisel, ja „vastuoluline tekst“, juhul kui tekstis esineb sama arv positiivseid ja negatiivseid sõnu. Kui tekstis on tegusõnavormid, mis viitavad soovitustele, siis annab programm hinnangule lisaks ka hoiatuse, et tegu võib olla soovitusega.

Peale arvamustes olevate hinnangute ja hoiakute leidmise oleks hea leida täpsemat informatsiooni lausetes sisalduvate tüüpiliste teemade kaupa, mis ainete arvustamisega seonduvad. Korpuses esinevad kõige sagedamad teemad saaks jagada kategooriatesse „aine“ (sageduselt neljas sõna, esines üle 20 000 korra), „õppejõud“ (esines üle 11 000

korra), „loeng” (esines üle 8000 korra) või „eksam“ (üle 6700 korra). Teemade leidmisel saab uurida, mis sõnadega viidatakse mõne teema aspektile. Näiteks on nii ainel tervena kui eksamitel maht ning raskusaste, kuid eksamite puhul võib ka esineda arvamusi hindamise või õpitava materjali hulga kohta. Õppejõudude kohta väljendatakse arvamusi hoopis teistest aspektidest lähtudes, näiteks sõbralikkus või esitusoskus. Selliseid omadusi taheti leida öeldistäite-konstruksiooni ning omadusi väljendavate sõnade ja fraaside abil, kuid kahjuks ei jäänud selleks aega.

Et tulemusi inimlugejale pädevalt esitada, trükkib programm tekstifaili arvamuse numbri korpuses (a), selles esinevad tundesõnad koos tunnet väljendava märgiga (b), tekstile antud hinnangu (c) ning muud omadused (d), nt

- (a) See on arvamus number 45
Arvamussõnad, mis tekstis esinesid:
- (b) *korralik* +
mahukas -
- (c) *Vastuoluline* arvamus.
- (d) *Viidet (arvamustekstile)*, mida otsida, polnud.

Tulemuste kontrollimiseks leiti korpusest juhusliku valiku teel sada arvamusteksti, mille sõnadel ja fraasidel märgendati käsitsi polaarsus ja leiti teksti üldine arvamus. Märgendamine tehti mõneti ka korpusega tutvumiseks, enne kui automaatse süsteemi analüüsi teostas. Käsitsi märgendatud tekstid on kättesaadavad autorilt tingimusel, et Tartu Ülikooli Õppeosakond seda lubab. Käsitsi märgendatud tekstide tulemuste ning süsteemi leitud hinnangute tulemusi võrreldi.

Tekstides esines slängi või muudmoodi ebastandardset keelt, kuid enamjaolt olid need kirjakeelsed. Arvamused olid pigem lühikesed, alustades kahest-kolmest sõnast kuni viie-kuue lauseni.

Käsitsi märgendatud lausetest selgus, et hoiakuid sisaldavate tekstide arv oli küllaltki suur – 77%, ülejäänud jagunesid kas hoiakutega tekstideks, hoiakuteta tekstideks (otsesed soovitusel, faktid ainete korralduslike poole kohta) või ebasobilikeks tekstideks (tekstid, mis polnud eesti keeles või polnud ainete tagasisidega seotud).

Hoiakutega tekstid kasutasid tüüpiliselt subjektiivset väljendusviisi. Näiteks viisil „*aine ei meeldinud, õppejõud oli halb*“. Oli ka paar, mis seda otseselt ei kasutanud.

Hoiakuteta arvamused olid tüüpiliselt soovitusel stiilis „*tehke kodutööd, ja ärge unustage kordamisküsimusi vaadata*“, kuid esines ka näiteks subjektiivne kommentaar selle kohta, kuidas arvaja leiab, et oleks võinud ikka loengus rohkem sõna võtta.

Ebasobilikud tekstid kas polnudki otseselt tekstid, näiteks tähe- või kirjavahemärgijadad näiteks „-“ või „*shdg*“, või olid teises keeles. Selles valimis leidis peale eestikeelsete lausete ainult inglisekeelseid tekste. Muukeelsed tekstid olid samuti lühikesed ning üldiselt väljendasid arvamust.

Programmi koodi on märkuste ning vormistusega 145 rida. Selles on 11 funktsiooni ning kolm regulaaravaldist. Alustati tundesõnastiku sisselugemisega ning selle alusel emotsiooniga sõnade ja lemmade leidmisel. Hiljem lisati juurde regulaaravaldised eituse ning „ma arvan, et...“ fraasi variantide leidmiseks. Eitus kuni nelja sõna kaugusel pöörab lähedal oleva tundesõna tähenduse vastupidiseks. Näiteks lauses *see pole tore* on sõna *tore* negatiivse tähendusega, kuna selle ees on eitust näitav sõna *pole*. Lisaks üritati regulaaravaldisega leida soovitusi, et vältida neutraalsete lausete arvamusteks määramist. Suurimaks takistuseks koodi kirjutamisel olid probleemid täpitähtedega, täielikul korpuse analüüsil tekkisid probleemid Unicode-formaadiga. Mõningased probleemid olid ka tulemuste kättesaamisel, kuna programm küll suutis neid välja trükkida, kuid mitte erinevasse tekstifaili. See probleem lahendati väljatrüki käsuraal faili suunamisega.

Lisaks hoiaku analüüsi programmile kirjutati ka muud olulist. Valmistati saja juhusliku lause väljastamise programm, sagedamate lemmade ning sagedamate omadussõnade leidmise programm sõnastiku täiendamiseks ning võimalike teemade leidmiseks. Veel kirjutati programm, mis järjestaks sõnastikku tähestikuliselt ning hoiakute analüüsi tulemustest sagedamini esinevate arvamussõnade programm. Need programmid olid üsna lühikesed, ligikaudu 30-50 rida, ning kasutasid tihti samu koodijuppe nagu põhiprogramm.

2.1. Tulemused

Tulemused eksporditi *txt*-faili, tulemustest saadud informatsioon arvutati Microsoft Wordi „Otsi kõik“ funktsiooni abil.

Nagu näha tabelilt 1, on analüüsiks liiga lühikesi tekste korpuses programmi andmetel 5%, neutraalseid tekste 30%. Hoiakuid sisaldavaid tekste esines korpuses 65%, neist valdav enamik olid positiivsed (71%). Järgnesid negatiivsed arvamused, mis moodustasid peaaegu neljandiku hoiakutega arvamuste hulgast (22%). Kõige vähem oli vastuolulisi arvamusi, 7% hoiakutega arvamuste hulgas.

Kui võrrelda programmi määratletud tekste käsitsi märgendatutega, on märgata mõningaid erinevusi. Hoiakuid sisaldavate tekstide arv on palju väiksem, kui käsitsi märgendatud tekstide hulgas – üle kümne protsendi esineb neid vähem. Sellevõrra on neutraalseid lauseid rohkem. Käsitsi märgendatud tekstide jaotust kujutab tabel 2. Ilmselt tuleneb vahe sellest, et programmil polnud vahendeid kaudsemalt väljendatud emotsioonide ning lausekonstruktsioonide tuvastamiseks. See on siiski üsna kõrge arv – tekstide analüüs õigustas korpuse valiku, hoiakut väljendavate lausete arv on palju suurem kui tavalistes korpustes. Programmi saadud positiivsete (71%) ja negatiivsete arvamuste (22%) jaotused hoiakuid sisaldavate tekstide hulgas on väga sarnased käsitsi märgendatutega (positiivseid arvamusi 74%, negatiivseid 22%). Suurim vahe tekkis vastuoluliste arvamuste jaotuses – programmi analüüsida järgi on neid 7%, käsitsi märgendades peaaegu poole vähem, 4%.

Tabel 1. Rakenduse analüüsitud arvamustekstide arv ja jaotus, kõigi tekstide arv ja jaotus.

	Arvamustekstide arv	Osakaal arvamustega tekstidest	Osakaal kõigist tekstidest
Positiivne tekst	11384	71%	46%
Negatiivne	3594	22%	15%
Vastuoluline	1149	7%	5%
Kokku	16127	100%	65%
	Kõigi tekstide arv		
Neutraalne	7475		30%
Ebasobilik	1160		5%
Kokku	24762		100%

Tabel 2. Märgeandatud arvamustekstide arv ja jaotus, kõigi tekstide arv ja jaotus.

	Arvamustekstide arv	Osakaal arvamustega tekstidest	Osakaal kõigist tekstidest
Positiivne tekst	57	74%	57%
Negatiivne	17	22%	17%
Vastuoluline	3	4%	3%
Kokku	77	100%	77%
	Kõigi tekstide arv		
Neutraalne	16		16%
Ebasobilik	7		7%
Kokku	100		100%

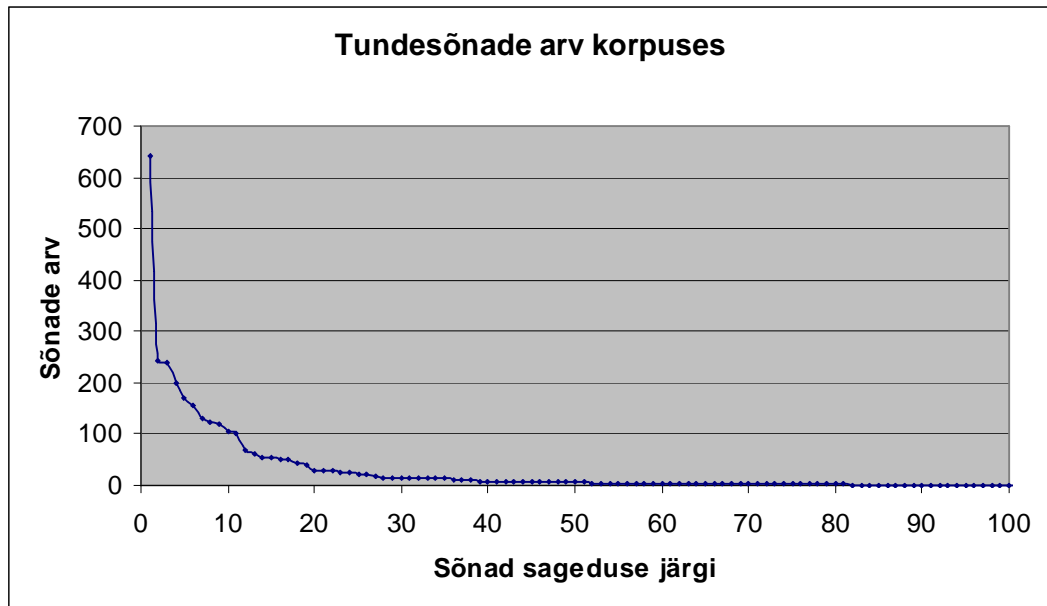
„Ma arvan, et...“-tüüpi lauset esines korpuses vaid kuus korda. Mõnel esinemiskorral ei väljendanud see otsest arvamust (kuigi väljendati kaudset). Kahjuks ei suutnud regulaaravaldis alati tervet järgnevat fraasi välja tuua, nii et otsisin seda teksti esinemisnumbri järgi. Ilmselt on see küll viis, mille abil saab arvamusi otsida, kuid neid ei leia eriti palju. Viitena võiks kasutada hoopis erinevaid arvamustega seotud sõnatüvesid, näiteks „*arvama*“, „*arvamus*“, „*tundma*“, „*aru saama*“ jne ning lihtsamaid väljendeid nagu „*minu meelest*“, „*minu arust*“.

Kuna soovitusi väljendavaid konstruktsioone võib tekstis olla palju, ilma et nende olemasolu pruugiks otseselt soovitusi väljendada, ei pidanud programm märkima nende olemasolul teksti automaatselt neutraalseks. Soovitusi leidmine osutus aga programmile üle jõu käivaks ülesandeks. Regulaaravaldis soovitusi leidmiseks oli valesti koostatud või tuli korpuselt vale informatsioon.

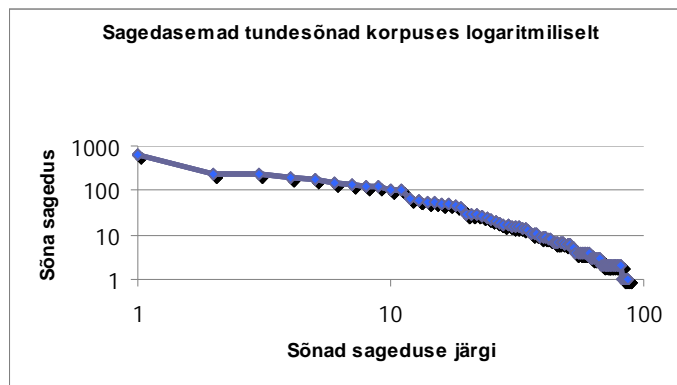
Kui aega oleks rohkem olnud, oleks leitud ka tingivas kõneviisis lauseid, et need välja tuua, ja leida öeldistäidete ning olemasolulausete abil peamiste teemadega seonduvad arvamused. Neid keelelisi nähtusi ning soovitusi väljendamist tasuks hoiakute analüüsis edasi uurida.

Tundesõnu leiti analüüsi käigus märgatavalt vähem, kui omadussõnade sageduse loend korpuses näitas (ligi kümme korda vähem). Ilmselt oli palju arvamusi väljendavaid sõnu just väga lühikestes arvamustes. Tundesõnade arvu tulemuste hulgas võib näha graafikul 1.

Võrdluseks esitatakse ka tundesõnade sagedus logaritmiliselt võrreldud Zipfi seaduse graafikuga, need on graafikud 2 ja 3.

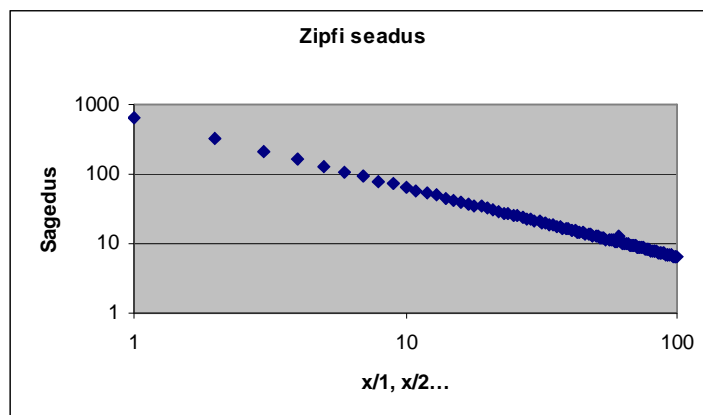


Graafik 1. Analüüsis leitud tundesõnade arv sageduse järgi.



Graafik 2. Sagedasemad tundesõnad korpuses logaritmiliselt.

Kui võrrelda sarnaseid langevaid graafikuid 2 ja 3, on näha erisusi. Sagedasemate tundesõnade graafiku joon langeb alguses aeglasemalt, lõpu poole aga palju kiiremini kui Zipfi seadust esindav graafik.



Graafik 3. Zipfi seadus.

KOKKUVÕTE

Hoiakute analüüsi uurimiseks ning rakenduste loomiseks on maailmas ning ka Eestis piisavalt huvi. Sotsiaalmeedias ja internetis kättesaadavate tekstide põhjal tasub teha nii keeleteaduslikke kui keeletehnoloogilisi katsetusi hoiakute ja hinnangute leidmiseks. On ka piisavalt takistusi ja uurimisküsimusi, mis vajaks lahendamist, alustades sõnastike koostamisest kuni keerulisemate lausetüüpide määramiseni.

Hoiakute analüüsi arendamiseks Eestis oleks tähtsaim korpuste täiendamine ning ka uute korpuste loomine, mida saaks märgendatud kujul hoiakute analüüsi süsteemide treenimiseks kasutada. See lihtsustaks ka uurimist, kuidas tunded keeles avalduvad, kuna keelenäited oleks olemas. Lisaks oleks vaja abitööriistu, nagu sõnastikud, selgelt formuleeritud keelereeglid ning eesti keele töötlemisega harjunud statistilist klassifitseerijat.

Erinevate tekstide analüüsimiseks on vaja erinevaid meetodeid ning erinevaid abitööriistu. Üldiselt jaotuvad hoiakute analüüsi süsteemid kaheks: reeglipärasteks ning statistilisteks. Reeglipärastes süsteemides kasutatakse rohkem uurijate koostatud sõnastikku ning keelereegleid, statistilises aga rohkem ennustusmudeleid ning sõnastiku automaatset kogumist. Lühemate ning normitud kirjakeelt sisaldavate tekstide puhul saab kasutada lihtsamaid lahendusi, pikemate ning kirjakeelest erinevate tekstide puhul tuleb konstrueerida keerulisemaid meetodeid.

Üliõpilaste tagasiside ainetele oli pigem lühem. Kuigi kasutati slängi, oli enamik tekstist kirjakeelne. Oli ka kirjavigu, kuid mitte nii palju, et see takistaks analüüsimist. Seepärast kasutati praktilises osa pigem lihtsamaid meetodeid, sõnastikuga otsimine ning teatud fraaside leidmine regulaaravaldisega. Kontrolliks märgendati enne praktilise osa programmi valmimist 100 juhuslikku lauset korpusest.

Praktilise osa programm andis informatsiooni nii arvamussõnade kohta, mis tekstis esinesid, kui ka hinnangu kogu tekstile ning otsis tekstidest „ma arvan, et...” fraasi erinevaid variante.

Selgus, et hoiakute analüüsi süsteem suutis suure osa tekste määrata hinnanguid väljendavateks tekstideks, kuid ei leidnud kõiki. See tulenes sellest, et programmil polnud abivahendeid kaudsete arvamuste leidmiseks ning lühemad arvamused jäeti analüüsimata. Hoiakuid väljendavate tekstide hulgas olid negatiivseks-positiivseks jagatud tekstid üsna samas osakaalus käsitsi märgendatutega.

Regulaaravaldistest kõige kasulikum oli eitus, mis märkis sõnale vastupidise hinnangu, kui sõna oli eituse lähiümbruses. Väga vähe oli tekste, kus esines fraasi „ma arvan, et...” variante, seega seda spetsiifilist fraasi ei soovitataks järgmiste uurimuste käigus viitena arvamusele kasutada. Soovitusi sisaldavaid tekste oli käsitsi märgendatute hulgas üsna mitu. Kahjuks ei funktsioneerinud regulaaravaldis, mis võimaldanuks soovitusi sisaldavaid tekste esile tuua.

Järgnevad uurimused võiks keskenduda konditsionaalis ning öeldistäidetes tunnete või hinnangute väljendamise reeglite formuleerimisele. See võimaldaks nii üldiselt meetodeid kui ka tundesõnade määramistäpsust rakendustes parandada.

KIRJANDUS

Dave jt = Dave, Kushal; Steve Lawrence; David M. Pennock 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of International Conference on World Wide Web (WWW-2003).

Eesti emotsionaalse kõne korpus. Eesti Keele Instituut. <http://peeter.eki.ee:5000/?lg=et>. Vaadatud 20.03.2014.

EKKT 2006–2010 = Eesti Keele Instituut. Uue meedia keele süntaktiline analüüs. Eesti keeletehnoloogiline tugi 2006–2010. <http://www.keeletehnoloogia.ee/projektid/eesti-keele-soltuvusgrammatika/suulise-keele-ja-uue-meedia-keele-tootlus/uue-meedia-keele-suntaktiline-analuus>. Vaadatud 18.05.2014.

Emotsioonidetektor. Eesti Keele Instituut. <http://peeter.eki.ee:5000/valence/>. Vaadatud 20.03.2014.

Google Prediction API. https://developers.google.com/prediction/docs/sentiment_analysis. Vaadatud 20.03.2014.

Himma, Marju 2014. Seadused võiks avaldada ka vene keeles. –Eesti Ringhääling. <http://uudised.err.ee/v/arvamus/963c5623-2b30-4864-b514-c63e7c7c1819>. Vaadatud 09.05.2014

KT 2011–2017 = Eesti Keele Instituut. Kõne ja teksti emotsionaalsuse statistilised mudelid. Eesti keeletehnoloogia 2011–2017. <http://www.keeletehnoloogia.ee/ekt-projektid/kone-ja-teksti-emotsionaalsuse-statistilised>. Vaadatud 20.03.2014.

Liu, Bing 2012. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Marran, Siim-Toomas 2012. Sentimentaalne analüüs eestikeelse peavoolumeedia veebiartiklite kommentaaride baasil. Bakalaureusetöö.

Microsoft Dynamics. <http://www.microsoft.com/en-us/dynamics/crm-social.aspx> Vaadatud 20.03.2014.

Nasukawa, Tetsuya; Jeonghee Yi 2003. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the K-CA P-03, 2nd International Conference on Knowledge Capture.

Pajupuu, Hille 2012. Emotsioonid – kõnetehnoloogia olevik ja tulevik. Keel ja Kirjandus, 8-9, 629–643.

Repustate Sentiment Analysis. <https://www.repustate.com/sentiment-analysis>. Vaadatud 09.05.2014.

Sentiment140. Alec Go, Richa Bhayani, Lei Huang. Stanford University. <http://www.sentiment140.com/>. Vaadatud 09.05.2014.

Tweet Sentiment Visualization. North Carolina State University. http://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/. Vaadatud 09.05.2014.

Vainik, Ene 2012. Kuidas määrata eesti keele sõnavara tundetoone? Eesti Rakenduslingvistika Ühingu Aastaraamat, 8, 257–274.

Vainik, Ene 2010. Kuidas õpetada kõnesüntesaatorile empaatiat? Emotsiooni automaatse tuvastuse võimalustest eestikeelses kirjalikus lauses sisalduva info põhjal. Eesti Rakenduslingvistika Ühingu aastaraamat, 6, 327–347.

Vainik, Ene 2002. Kas eestlased on "kuumaverelised"? Eestlaste rahvalikust emotsioonikategooriast. Emakeele Seltsi aastaraamat, 47, 63–86.

Wang, Sida; Christopher D. Manning 2012. ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, 90–94.

Weiss jt = Weiss, Sholom M; Nitin Indurkha; Tong Zhang 2010. Fundamentals of Predictive Text Mining. Springer. Texts in Computer Science, Volume 41.

Wilson jt = Wilson, Theresa; Janyce Wiebe; Paul Hoffmann 2005. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 347–354.

SENTIMENT ANALYSIS ON UNIVERSITY OF TARTU STUDENT FEEDBACK. SUMMARY

There has been a lot of interest towards *sentiment analysis* (also known as *opinion mining*) since the start of the 2000s, because of wider access to the Internet. Sentiment analysis is a method to automatically categorize the emotional content of different types of text. The driving force behind sentiment analysis is the large number of easily accessible anonymous opinions in the Internet, and commercial interest in building resources and applications using the method.

However, there hasn't been a lot of research into the subject of sentiment analysis in Estonian. Even work on expressing sentiments is sparse. Due to this there aren't many applications using the method.

The theoretical part of this thesis gives a thorough overview of opinion mining methods and their application for Estonian texts. The practical part describes the sentiment analysis of student feedback about university subjects collected by the University of Tartu. The analysis shows the estimated number of positive and negative opinion texts in the feedback. Most of the sentiment expressions were found and texts evaluated using a sentiment lexicon in conjunction with regular expressions.

As expected, the texts contain a large number of opinions. More than 65% of the corpus contains sentiments according to the computer program developed by the author. Manually analysed 100 random texts from the corpus were used to compare the results from the analysis. The percentages of negative and positive opinions were identical to the manual analysis, but a lower percentage of opinions were found by the program than manual analysis.

To advance sentiment analysis in Estonian, more research should be conducted into the use of the conditional and more text corpora should be made available to test analysis programs.

LISAD

Lisa 1. Tundeleksikon

Positiivsed sõnad	Negatiivsed sõnad
abivalmis	Arusaamatu
abivalmilt	arusaamatus
avameelne	arusaamatult
avameelselt	ebamoraalne
avardav	ebamäärane
aktuaalne	ebaselge
arendav	ebasoodne
armastusväärsus	ebasõbralik
aru saada	ebaõnnestuma
arukas	ebaõnnestumine
arukalt	edutu
arusaadav	edutult
arusaadavalt	eksimus
asjalik	eksitama
asjatundja	eksitus
avardav	halb
detailne	halvaendeline
detailsest	haltuura
efektiivne	hirmus
efektiivselt	hirmuäratav
edenema	hoolimatu
edukalt	hoolimatult
edukas	hoolimatus
edusamm	hukkamõistetav
eluline	hull
enesekindel	hulluma
enesekindlalt	hullumeelne
enesekindlus	hullus
entusiastlik	huvipuudus
ergutav	häbistama
esilekerkiv	häda
esmajärguline	häda ja viletsus
esmaklassiline	häda ja õnnetus
haaravalt	hädavaevu
hariv	hädine
hea	häirima
heakvaliteediline	häiriv
heameelega	hämar
heaolu	hägune
heasoovlikkus	igav
heatahtlik	igavapoolne
hoolitsema	igavavõitu
hubane	igavlema
huvipakkuv	igavus
huvitatud	ilmetu
huvitav	ilmetult
huvitekitav	inetu
huviäratav	inetult

hõlpsalt	jube
hõlpsasti	juhm
häirimatu	jõle
hüva	jäle
hüvang	jälk
ilmekas	kaebama
ilmekalt	kahju
imelihtsalt	kahjulik
imeline	kammitsema
imeliselt	kartma
imetore	katsumus
julgesti	keeruline
jõukohane	keerutama
jõupärane	kehv
kaasahaarav	kibekiire
kaasakiskuvalt	kohutav
kasulik	kole
kaval	konflikt
kavalalt	korratu
keerutamata	kuivalt
kerge	kuri
kiiduväärne	kurjalt
kiiduväärt	kurnama
kiitma	kurtma
kiitusväärne	käegalöömine
kodune	kärsitu
koduselt	kärsitus
korralik	külm
korralikult	külmalt
korrektselt	labane
kvaliteetne	laiduväärne
kõlbab	laiduväärt
kõrgeklassiline	laisklema
kõrgetasemeline	laitma
käkitegu	lohakas
käsitama	lohakalt
kõitev	luhtuma
kõitvalt	läbikukkuma
laabuma	läbikukkunud
ladus	lünklik
ladusalt	mahukas
laiendav	maniakk
laitmatu	marutõbine
lahe	metsik
lahedalt	miinus
lahedasti	monotoonne
lahke	monotoonselt
lepitama	mula
lihtne	mulisema
lihtsasti	murelik
loogiline	mõistmatu
ludinal	mõnitama
läbitav	mõttetu
lõbus	mõla
maitsekas	mõlisema

meelikõitev	naeruväärne
meelihaarav	napilt
meeldima	negatiivne
meeldiv	nihu
meeleldi	nilbe
meelsasti	nilbelt
meeltmööda	nuhtlema
meisterlik	nuhtlus
mugav	nurjuma
muretu	nõrk
mõikama	nõutu
mõistlik	närune
mõistma	nördima
mõistuspärane	ogar
mõjus	ohtlik
mõnus	ohtlikult
mõnusasti	ohustama
märkimisväärne	paha
oivaline	passiivne
oluline	pealiskaudne
omandatud	peletama
osav	peletav
osavalt	perversne
otsekohene	perversselt
paeluv	petma
paeluvalt	pidurdama
parim	pinnapealne
pehmesüdameline	puuduma
pluss	puudus
pooldama	põlgama
positiivne	pärssima
praktiline	pöörane
priima	raev
põhjalik	raevukas
põnev	rahutu
põnevalt	raske
püsiv	raskus
rahulik	rist ja viletsus
rahustav	ropp
rõõmus	ropult
rõõmustama	rumal
saavutama	rumalalt
saavutatav	rutakalt
selge	ruutama
silmapaistev	rõve
silmatorfav	rõvedalt
sirgjooneliselt	räige
sobilik	räigelt
sobiv	rämps
sobivalt	ränkraske
soe	salgama
soliidne	segama
soodne	segane
soojalt	segav
soojasüdameline	selguseta

soovitama	selgusetu
soovitativ	siivutu
sorav	surm
soravalt	surema
sujuma	surmigav
sundimatu	sõge
suurepärane	sõgedalt
sõbralik	takistama
sõbralikkus	takistus
südamlik	taluma
taibukas	taunitav
taipama	taunima
takistusteta	tuim
tasakaalus	tuimalt
tasakaalukas	tulemusteta
tasuma	tulutult
tehtav	tõrjuv
toetama	tõrjuma
toetav	täbar
tore	tüli
tugev	tülitsema
tulemuslik	tüütu
tulus	ummik
tunnustama	unine
tõhus	vajakajäämine
tõhusalt	vaenulik
tõrgeteta	vaev
tähtis	vaevaliselt
vaevata	vale
vaheldusrikas	valelik
vaimustav	valetama
vajalik	valus
valima	vastutustundetu
vastutulelik	viga
veatu	vigane
veatult	vihane
vigadeta	vihasama
virgutav	vilets
võluv	vulgaarne
võrratu	võigas
võtma	võimatu
väljapaistev	õnnetu
värske	õnnetus
väärt	õudne
väärtuslik	äparduma
õdus	äpardus
õnnelik	ärevil
õnnestunud	ärevus
üksikasjalik	ükskõikne
ülilahe	üksluine
ülitore	ülbe
	ülbitseja

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina

_____ (autori nimi)

(sünnikuupäev:

)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

_____,
(lõputöö pealkiri)

mille

juhendaja

on

_____,
(juhendaja nimi)

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus/Tallinnas/Narvas/Pärnus/Viljandis, _____ (kuupäev)